

Diamonds price prediction

Done by:
Renad Albishri
Eman Alshehri



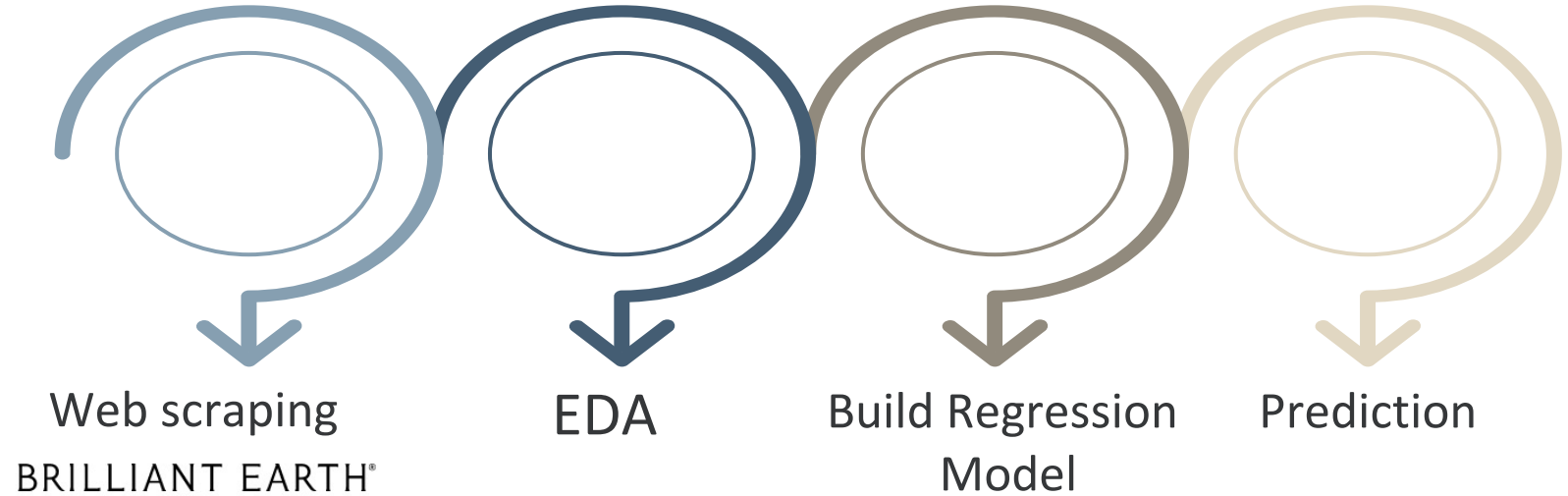
Overview

In this project, we are working on a dataset that consists of information about the diamonds and features that price depend on, such as Clarity, Color ,etc.

Goals

- Which variables are significant in predicting the price of a diamonds
- Build a model to predict the diamonds price.
- Choose the model that give us the best predict.

Methodology



Ends Soon! Complimentary Diamond Necklace with Purchase Over \$1,000

800.691.0952



Virtual Appointment

BRILLIANT EARTH®



Engagement

Wedding

Diamonds

Gemstones

Jewelry

About

Search

Diamonds / Search for Diamonds

Search for Diamonds

Natural

Lab



Need help finding the perfect diamond? [Start with Our Diamond Quiz](#)

SHAPE ?



PRICE ?



\$450

\$463,580

CARAT ?



0.30

12.05

CUT ?



Fair

Good

Very Good

Ideal

Super Ideal



CHAT

COLOR ?



J

I

H

G

F

E

D

CLARITY ?



SI2

SI1

VS2

VS1

VVS2

VVS1

IF

FL

Dataset

The dataset used in this project was obtained by scraping data from brilliant earth(online website) that offers one of the largest collections of diamonds available for sale.

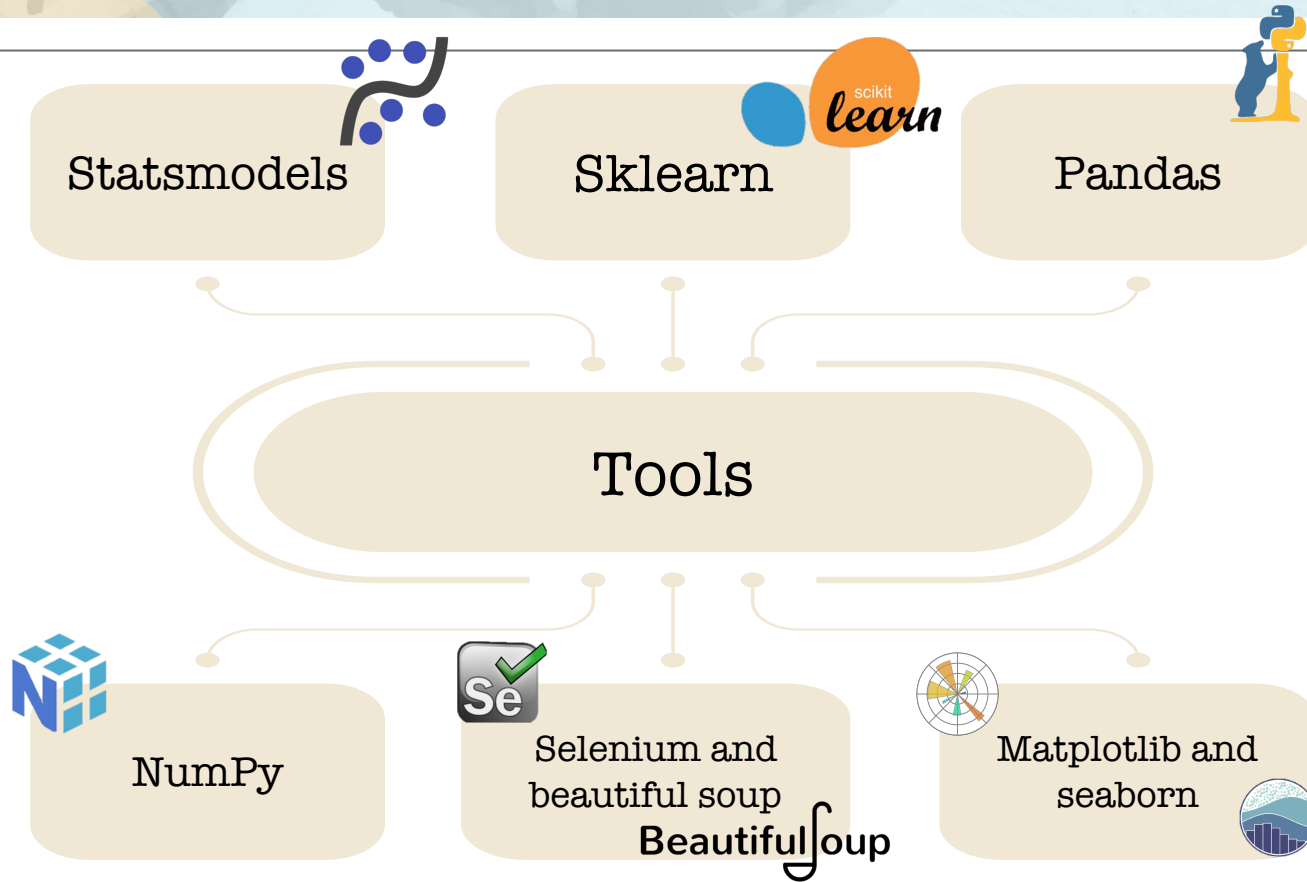
Dataset contains :

119,000 rows

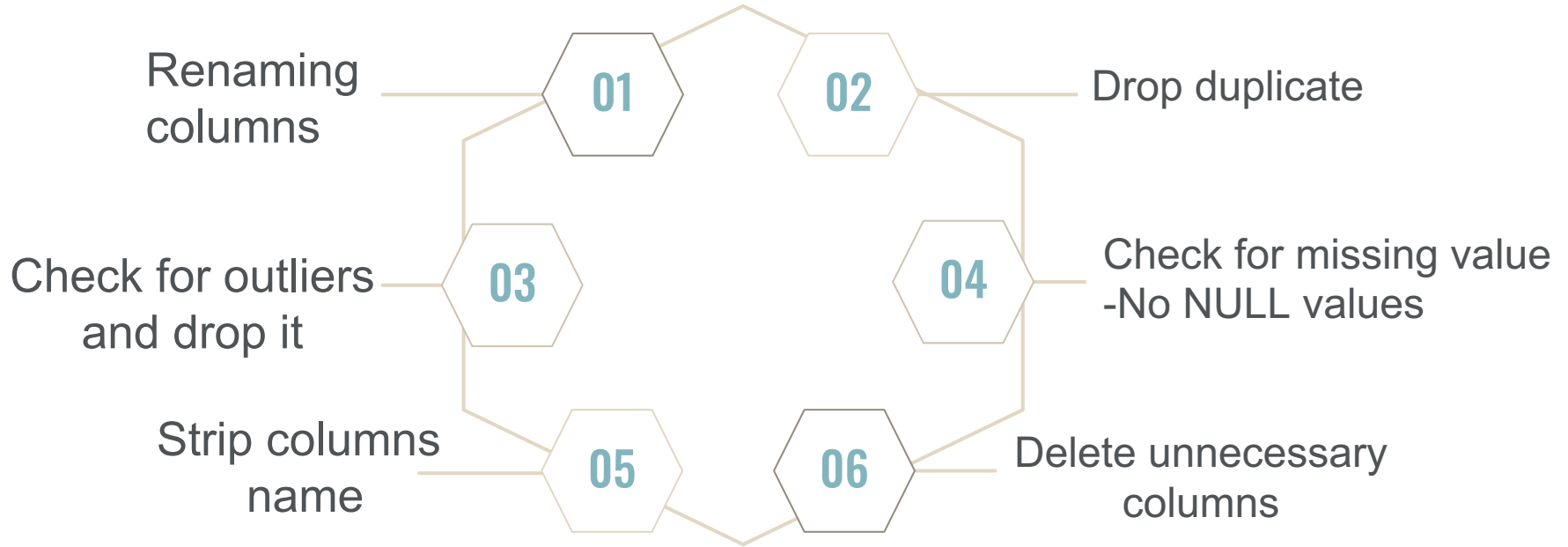
10 features

Id, URL , Shape , Carat , Cut , Color , Report , Type , LW , Price

It's a great dataset for machine learning and working with data analysis and visualization.

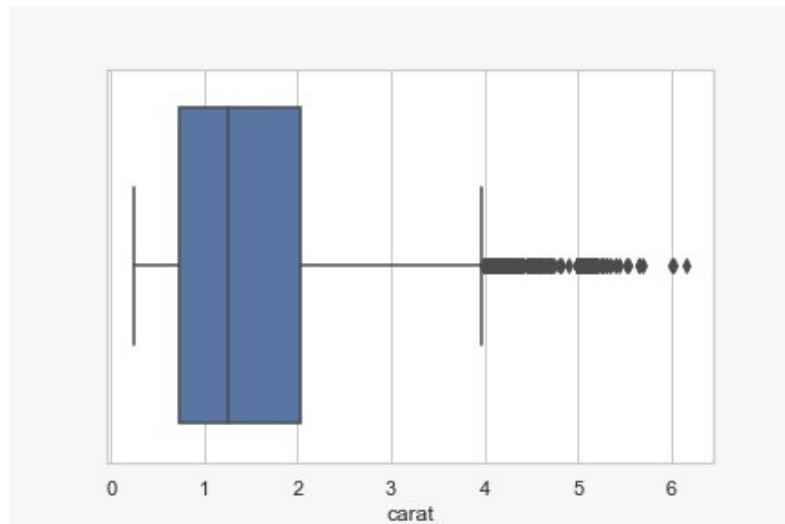


EDA

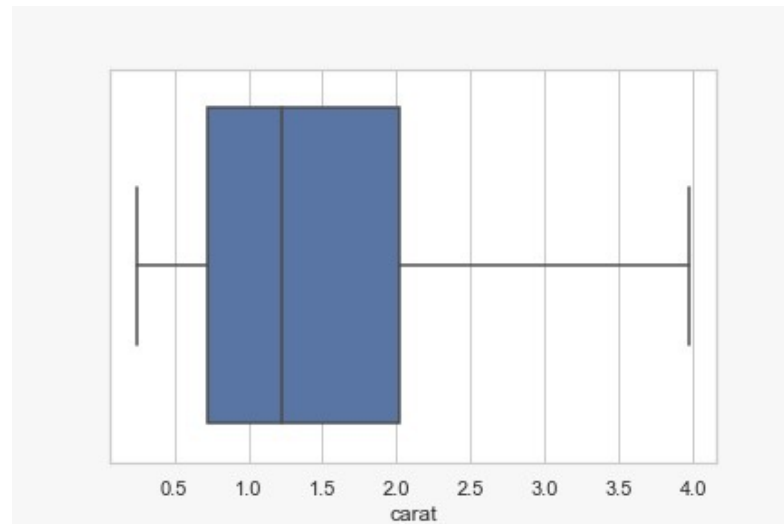


Outlier

before



After

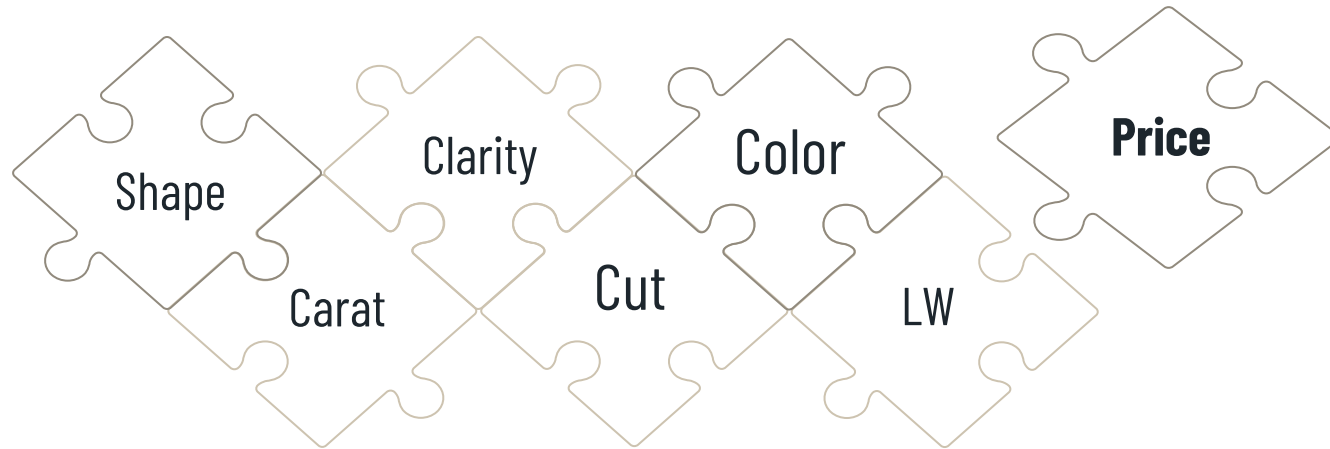


EDA cont....

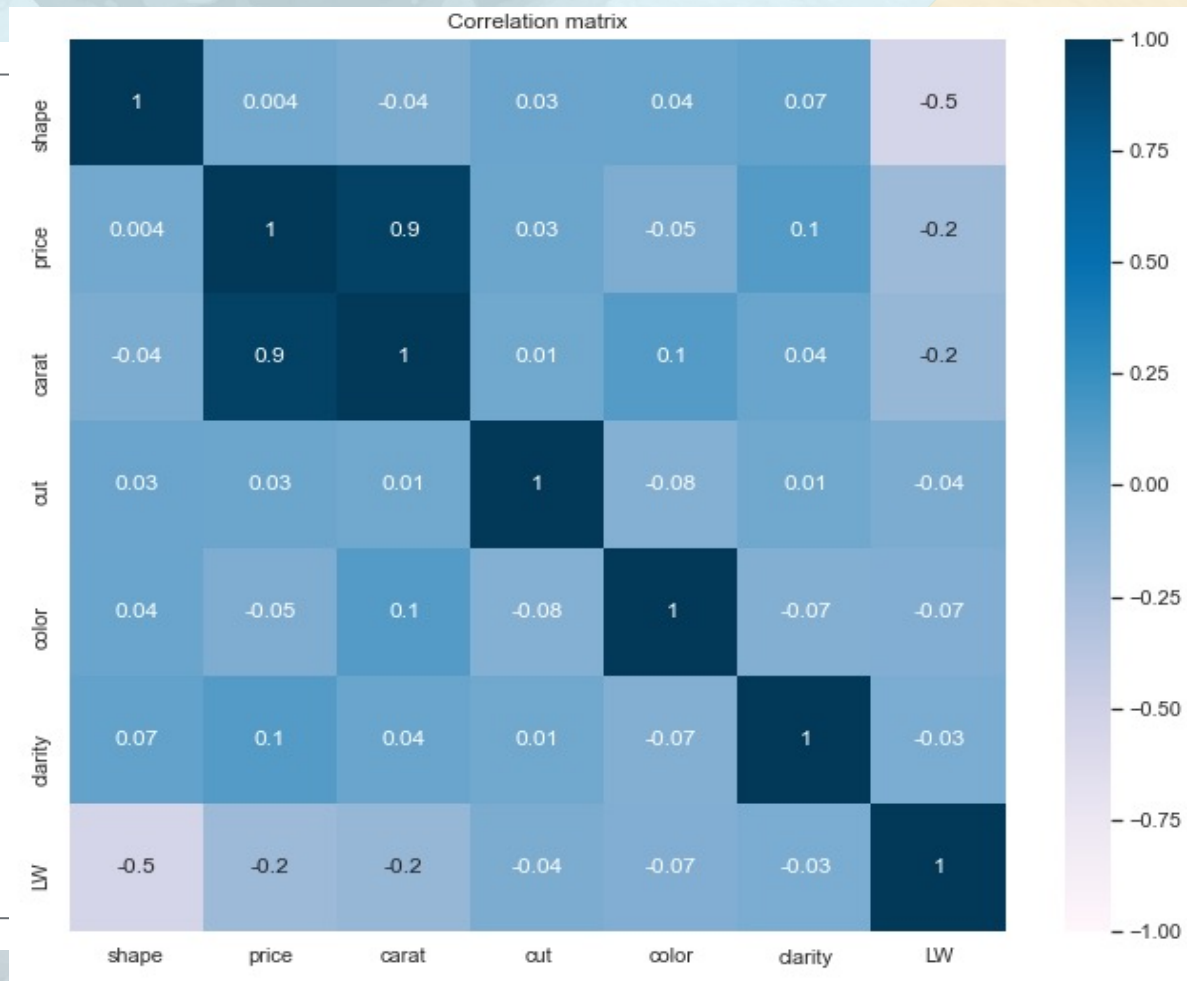
After cleaning the data, the dataset becomes:

65,000 rows

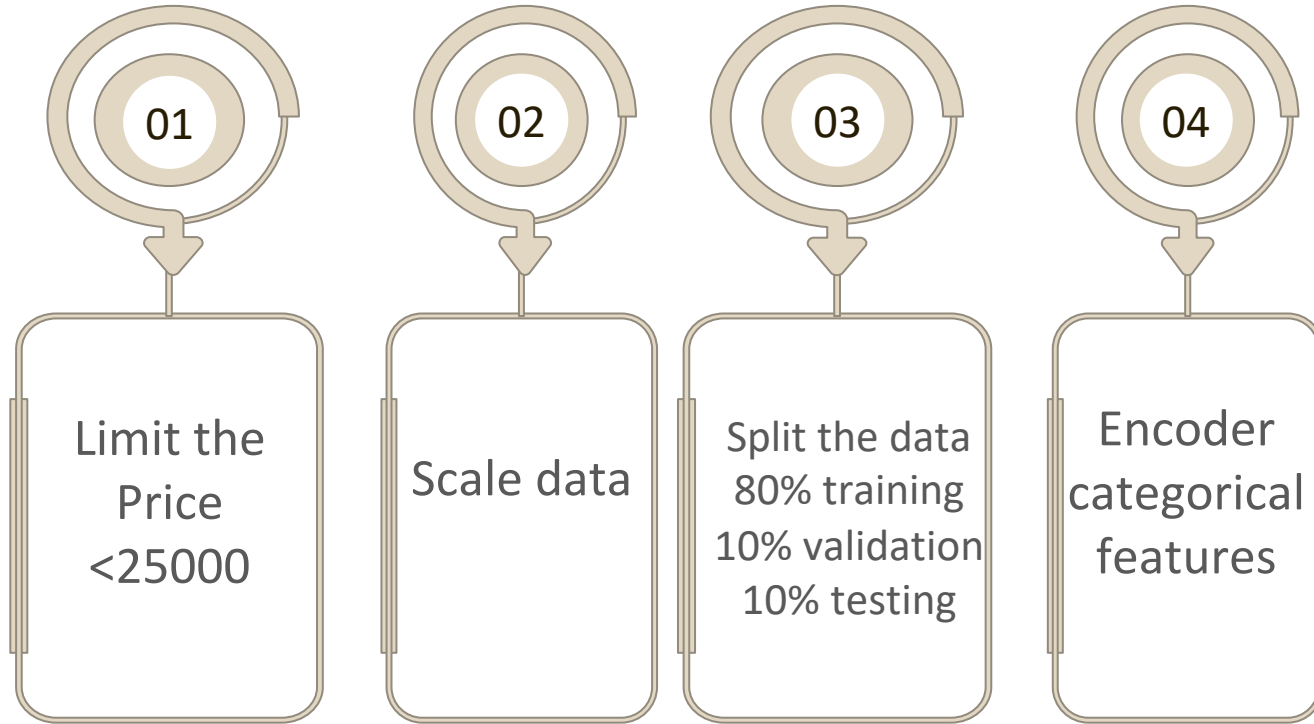
7 features



Features and target correlation



Regression and feature engineering



Feature engineering

Before

Linear Regression val R^2 : 0.854
Ridge Regression val R^2 : 0.854
Degree 2 polynomial regression val R^2 : 0.947
Degree 3 polynomial regression val R^2 : 0.958
Degree 4 polynomial regression val R^2 : 0.964
Degree 5 polynomial regression val R^2 : 0.967
Degree 6 polynomial regression val R^2 : 0.972
Degree 7 polynomial regression val R^2 : 0.952
Lasso Regression val R^2 : 0.854
ElasticNet Regression val R^2 : 0.854

After

Linear Regression val R^2 : 0.904
Ridge Regression val R^2 : 0.904
Degree 2 polynomial regression val R^2 : 0.969
Degree 3 polynomial regression val R^2 : 0.975
Degree 4 polynomial regression val R^2 : 0.979
Degree 5 polynomial regression val R^2 : 0.981
Degree 6 polynomial regression val R^2 : 0.979
Degree 7 polynomial regression val R^2 : 0.976
Lasso Regression val R^2 : 0.875
ElasticNet Regression val R^2 : 0.904

Predicted models

	R ² training	R ² validation
Linear Regression	0.905	0.904
Polynomial Regression degree 4	0.977	0.979
Polynomial Regression degree 5	0.980	0.981
Polynomial Regression degree 6	0.984	0.979
Lasso Regression	0.875	0.875
Ridge Regression	0.905	0.904
Elastic Net	0.905	0.904

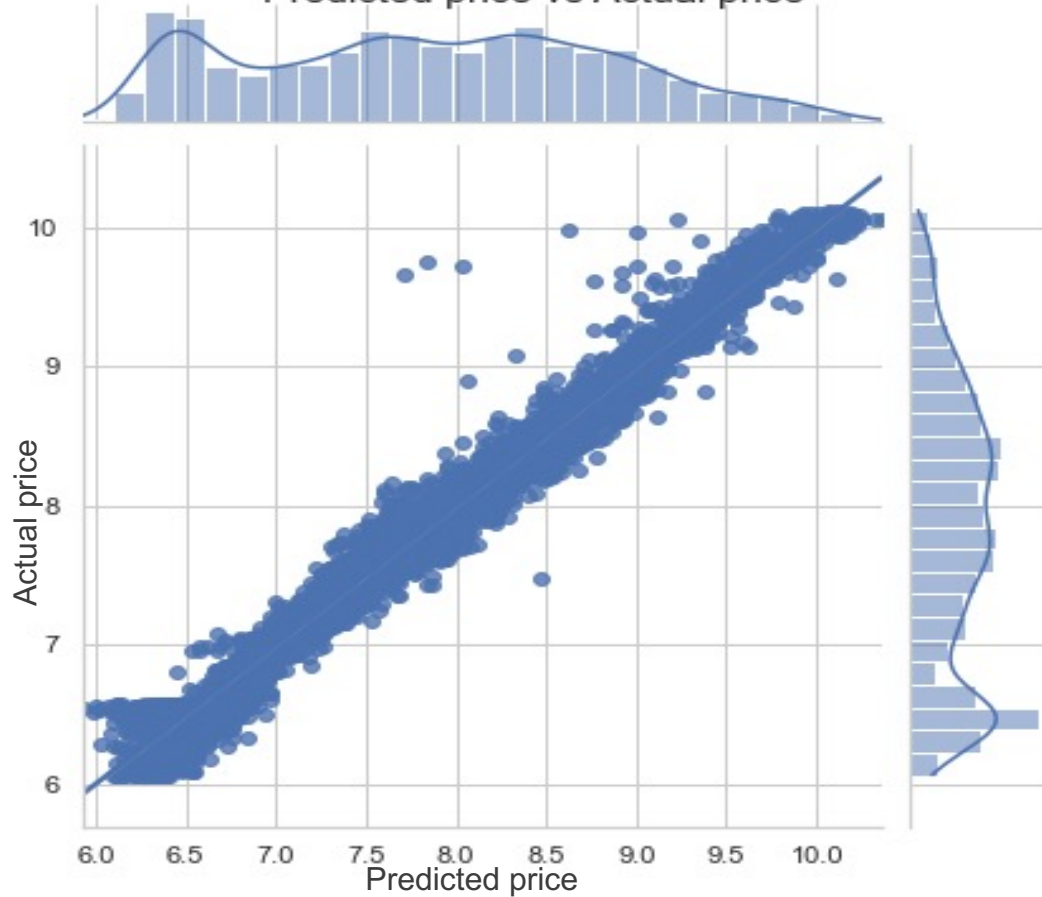
Predicted models

	R ² training	R ² validation
Linear Regression	0.905	0.904
Polynomial Regression degree 4	0.977	0.979
Polynomial Regression degree 5	0.980	0.981
Polynomial Regression degree 6	0.984	0.979
Lasso Regression	0.875	0.875
Ridge Regression	0.905	0.904
Elastic Net	0.905	0.904

Selected model

	R² test	RMSE
Polynomial Regression degree 5	0.9807	0.1413

Predicted price vs Actual price



The result of the best model
(Polynomial Regression $d=5$)

Conclusion

- Most significant feature in predicting the price of a diamonds is **Carat**.
- The least features affect in predicting the price of a diamonds are **Color, LW, and shape**.
- Selection model to predict diamonds price is **Polynomial Regression(degree 5)**